



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1460
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445

7590 12/21/2006
William E Lewis
Ryan Mason & Lewis LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER

HILLERY, NATHAN

ART UNIT	PAPER NUMBER
----------	--------------

2176

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	12/21/2006	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No. 09/703,174	Applicant(s) AGGARWAL ET AL.	
	Examiner Nathan Hillery	Art Unit 2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 02 October 2006.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-27 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-27 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This action is responsive to communications: Amendment filed on 10/2/06.
2. Claims 1 – 27 are pending in the case. Claims 1, 10, and 19 are independent.

Claim Rejections - 35 USC § 112

3. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

4. Claims 1 – 27 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

5. **Regarding independent claims 1, 10, and 19**, it is unclear what Applicant means by **without assuming an initial model of a link structure**. The specification recites, “not assuming a specific model for web linkage structure” (Specification, p 4, line 24). Consequently, the metes and bounds of the claim are unclear. Thus, the phrase **without assuming an initial model of a link structure** will be interpreted as “without assuming a specific model for web linkage structure” for analysis in the rejection of claims under 35 USC 102 and 103 below.

6. **Regarding dependent claims 2 – 9, 11 – 18 and 20 – 27**, the claims are rejected for fully incorporate the deficiencies of the base claim(s) from which they depend.

Claim Rejections - 35 USC § 101

7. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

8. Claims 1 – 27 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. Claims 1 – 27 have no practical application as claimed because there is no physical transformation and no production of a concrete, useful and tangible result.

a. The result of the claimed invention remains in the abstract and is not made available to the user; thus it is not tangible.

b. The claims appear to be in the preliminary stages and fall short of the disclosed practical utility. In other words, the claims fail to fulfill and/or reflect the specific, substantial, and credible utility sought by the disclosed invention, and thus do not produce a useful result.

c. The input retrieved and collected by the invention appear to be subjectively analyzed with no reliable, assured result being output, and thus does not produce a concrete result.

9. Consequently, the claims are nonstatutory. The claims simply recite retrieving and collecting data and/or information with no concrete, useful, tangible result.

10. Further, to expedite a complete examination of the instant application the claims rejected under 35 U.S.C. 101 (nonstatutory) above are further rejected as set forth below in anticipation of applicant amending these claims to place them within the four statutory categories of invention.

Claim Rejections - 35 USC § 102

11. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

12. Claims 1 – 8, 10 – 17 and 19 – 26 are rejected under 35 U.S.C. 102(b) as being anticipated by Chakrabarti et al. (Focused Crawling: A New Approach to Topic-specific Web Resource Discovery) [as cited by applicant].

13. **Regarding independent claim 1**, Chakrabarti et al. teach that *keyword search is used to locate an initial set of pages (using a giant crawl and index)* (p 6, section 2.2, last paragraph), which meet the limitation of **initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without assuming an initial model of a link structure**. It should be noted that the *keyword search* of Chakrabati et al. is equivalent to the claimed **user-defined predicate**.

14. Chakrabarti et al. teach that while fetching a document, the above formulation is used to find the leaf node with the highest probability. If some ancestor has been marked good we allow future visitation of URLs found on the document, otherwise the crawl is pruned there (p 9, section Hard focus rule), which meet the limitation of **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed and using the collected statistical information to automatically determine further document retrieval**

Art Unit: 2176

operations, since the probabilities are calculated to find the “best” leaf node, the ancestors are analyzed to determine if they are good, and then based on that finding future visitations are allowed (p 9, section Hard focus rule). It should be noted that the *probabilities* of Chakrabarti et al. are equivalent to the claimed **statistical information**.

15. Chakrabarti et al. teach that a focused crawler is an example-driven automatic porthole-generator. We feel that the ability to focus on a topical subgraph of the Web, as in this paper, together with the ability to browse communities within that subgraph, will lead to significantly improved Web resource discovery (p 3, last paragraph before Section 2), which meet the limitation of **wherein the statistical information-using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation**. It should be noted that the *porthole*, which is a *subgraph of the Web*, generated by the *focused crawler* of Chakrabarti et al. is equivalent to the claimed **link structure** that is learned. It should further be noted that the generation of a porthole or specialized link structure (p 20, last paragraph) is equivalent to the claimed **learning a link structure**.

16. **Regarding dependent claim 2**, Chakrabarti et al. teach that Query construction is not a one-time investment, because as pages on the topic are discovered, their additional vocabulary must be folded in manually into the query for continued discovery (p 7, lines 4 – 6), which meet the limitation of **the user-defined predicate specifies content associated with a document**. It should be noted that the *additional*

Art Unit: 2176

vocabulary of pages on the topic of Chakrabati et al. is equivalent to the claimed **content associated with a document**.

17. **Regarding dependent claims 3 and 4**, Chakrabarti et al. teach that pages that are examples associated with a topic can be preprocessed as desired by the system. The user's interest is characterized by a subset of topics that is marked good. No good topic is an ancestor of another good topic. Ancestors of good topics are called path topics. Given a Web page, a measure of its relevance must be specified to the system (p 8, lines 9 – 14), which meet the limitation of **the statistical information collection step uses content of the one or more retrieved documents** and that **the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents**, since a determination is made about the ancestors and preprocessed pages are used, which are equivalent to the claimed **one or more retrieved documents**. It should be noted that the *topic* of Chakrabarti et al. is equivalent to the claimed **content** and **predicate**.

18. **Regarding dependent claims 5 and 6**, Chakrabarti et al. teach that we have presented evidence in this section that focused crawling is capable of steadily collecting relevant resources and identifying popular, high-content sites from the crawl, as well as regions of high relevance, to guide itself. It is robust to different starting conditions, and finds good resources that are quite far from its starting point. In comparison, standard crawlers get quickly lost in the noise, even when starting from the same URLs (p 20,

Art Unit: 2176

Section 4.8 and p 18, Figure 9), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate**, and that **the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate than would otherwise occur with respect to document retrieval operations that are not directed using the collected statistical information**, since the focused crawling of Chakrabati et al. utilizes statistical information (p 3) and compares their crawler to other crawlers and outlines the other's shortcomings (Fig 9).

19. **Regarding dependent claim 7**, Chakrabarti et al. teach that multiple citations from a single document are likely to cite semantically related documents as well. This is why the distiller is used to identify pages with large numbers of links to relevant pages (p 8, last paragraph), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate**. It should be noted that the semantically related documents of Chakrabarti et al. is equivalent to the **claimed documents which are linked to by other documents which also satisfy the predicate**

Art Unit: 2176

20. **Regarding dependent claim 8**, Chakrabarti et al. teach that we describe a Focused Crawler, which seeks, acquires, indexes, and maintains pages on a specific set of topics that represent a relatively narrow segment of the Web. Thus, Web content can be managed by a distributed team of focused crawlers, each specializing in one or a few topics (p 2, fourth paragraph), which meet the limitation of **the information network is the World Wide Web and a document is a web page**.

21. **Regarding claims 10 – 17 and 19 – 26**, the claims incorporate substantially similar subject matter as claims 1 – 8, and are rejected along the same rationale.

Claim Rejections - 35 USC § 103

22. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

23. Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. as applied to claims 1 – 8, 10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by applicant] later referenced as Ch2 et al.

24. **Regarding dependent claim 9**, Chakrabati et al. do not explicitly teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

Art Unit: 2176

25. Ch2 et al. teach that other strategies are also known, such as, if the URL is of the form `http://host /path`, then the crawler may truncate components of path and try to fetch these URL's. If links could be traversed backward, e.g. using metadata at the server, the crawler may also fetch pages that point to the page being 'expanded.' (p 382, Column 1, lines 29 – 37), which meet the limitation of **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.**

26. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teachings of Chakrabarti et al. with that of Ch2 et al. because such a combination would provide the users of Chakrabarti et al. with teachings of *the architecture of a hypertext resource discovery system using a relational database* (p 375, Column 1, lines 1 & 2).

27. **Regarding claims 10 – 27**, the claims incorporate substantially similar subject matter as claims 1 – 9, and are rejected along the same rationale.

Response to Arguments

28. Applicant's arguments filed 10/2/06 have been fully considered but they are not persuasive.

29. Applicant argues that claims 1 – 27 are statutory because the retrieval operations can be performed outside of a computing device (p 9, second full paragraph).

The Office disagrees.

It should be noted that the test for statutory subject matter in a claim does not depend on whether or not steps can be performed outside of a computing device or combinations thereof. The test for statutory subject matter is whether there is a practical application of a judicial exception. To have a practical application of judicial exception present in the claims, the claims must provide a physical transformation or produce a concrete, useful and tangible result.

The claims fail to physically transform an object. The claims also fail to produce a concrete, useful and tangible result. If the result alleged by applicant is document retrieval, the result must be tangible, i.e. made available to the user in the real world.

30. Applicant argues that Chakrabarti et al. do not teach or suggest **wherein the initial document retrieval operation is performed without assuming an initial model of a link structure**, which has been interpreted as wherein the initial document retrieval operation is performed without assuming a specific model for web linkage structure (Specification, p 4, line 24) because Chakrabarti initiates crawling with a linkage sociology (p 9, last paragraph – p 10).

The Office disagrees.

It should be noted that, by Applicant's own admission, Chakrabarti refers to "*discovering* linkage sociology," (p 10, first line). Therefore, Chakrabarti does not assume a specific model of web linkage; Chakrabarti actually "discovers" the linkage sociology as it goes.

Art Unit: 2176

Using further evidence provided by applicant's own admission, Chakrabarti teaches, "The system starts by visiting all pages in D(C)" (p 10, first full paragraph). This citation of Chakrabarti further proves that Chakrabarti initially visits or retrieves all pages indiscriminately.

Then, Chakrabarti goes on to teach that [i]n each step, the system can inspect its current set V of visited pages and then choose to visit an unvisited page from the crawl frontier, corresponding to a hyperlink on one or more visited pages (p 10, first full paragraph). Finally, Chakrabarti teaches that [i]nformally, the goal is to visit as many relevant pages and as few irrelevant pages as possible, i.e., to maximize average relevance (p 10, first full paragraph).

Thus, initially all pages are visited then the system learns the linkage sociology by analyzing the visited pages with the ultimate goal of maximizing relevance by deciding which pages are relevant.

Conclusion

31. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not

Art Unit: 2176

mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillery whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on (571) 272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

H Herndon

Heather R. Herndon
Supervisory Patent Examiner
Technology Center 2100

NH